Crystallization and preliminary X-ray studies of snake gourd lectin: homology with type II ribosome-inactivating proteins

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray studies of snake gourd lectin: homology with type II ribosome-inactivating proteins

N. Manoj,[a] A. Arockia Jeyaprakash,[a] J. V. Pratap,[a] Sneha Sudha Komath,[b] Roopa Kenoth,[b] Musti. J. Swamy[b] and M. Vijayan[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India, and [b]School of Chemistry, University of Hyderabad, Hyderabad 500 046, India

Correspondence e-mail: mv@mbu.iisc.ernet.in

Received 17 January 2001
Accepted 14 March 2001

The lectin from the seeds of snake gourd (*Trichosanthes anguina*) has been crystallized in two forms using the hanging-drop method. Both the forms are hexagonal, with the asymmetric unit containing one subunit consisting of two polypeptide chains linked through disulfide bridges. Intensity data from one of the forms were collected at room temperature as well as at low temperature to 3 Å resolution. Molecular-replacement studies indicate that the lectin is homologous to type II ribosome-inactivating proteins. Partial refinement confirms this conclusion.

1. Introduction

Lectins are carbohydrate-binding proteins that specifically recognize diverse sugar structures and mediate a variety of biological processes such as cell–cell and host–pathogen interactions, serum glycoprotein turnover and innate immune responses (Lis & Sharon, 1998; Vijayan & Chandra, 1999). Of them, plant lectins are the most extensively and thoroughly characterized (Lis & Sharon, 1998; Loris *et al.*, 1998; Bouckaert *et al.*, 1999). The different families, well characterized structurally, include one involving plant toxins. Ricin (Rutenber *et al.*, 1991), abrin (Tahirov *et al.*, 1995) and mistletoe lectin (Krauspenhaar *et al.*, 1999) are lectins of known structure belonging to this family. Each of them typically consists of two polypeptide chains, each with a well characterized structure and often referred to as *A* and *B* domains, connected by disulfide bridges. They are often called type II ribosome-inactivating proteins (type II RIPs). The *A* chain is the toxic component and inactivates ribosome. The *B* chain is the lectin component and is presumably involved in recognition prior to entry into the cell. Type I ribosome inactivating proteins (type I RIPs) such as gelonin (Sairam *et al.*, 1993; Singh *et al.*, 1992), trichosanthin (Xiong *et al.*, 1994) and momorcharin (Ren *et al.*, 1994) consist only of the *A* domain, while there are lectins such as amaranthin (Transue *et al.*, 1997) which consist only of the *B* domain. The latter obviously do not have any ribosome-inactivating activity. However, as indicated earlier, type II RIPs combine toxic and lectin activities on account of the simultaneous presence of the *A* and *B* domains.

Macromolecular and carbohydrate-binding properties of several seed lectins from the Cucurbitaceae family have been investigated (Falasca *et al.*, 1989; Bostwick *et al.*, 1994; Komath & Swamy, 1998; Padma *et al.*, 1999). Detailed structural characterization of any of these has not been possible on account of, amongst other things, the non-availability of their complete amino-acid sequence data. Various lines of evidence suggest that they are structurally homologous to type II RIPs. However, they do not inactivate ribosomes or do so only weakly (Wang & Ng, 1998; Li *et al.*, 2000). The X-ray analysis of the seed lectin from *T. anguina*, a member of the Cucurbitaceae family commonly known as snake gourd, has been undertaken to elucidate the structural basis of this behaviour and also to explore the structure–function relationship in this family of lectins. Snake gourd seed lectin (SGSL) is glycosylated and has an acidic pI of 5.0. The lectin displays haemagglutinating activity and shows a high affinity for methyl-β-D-galactose at the monosaccharide level and for lactose at the disaccharide level (Komath & Swamy, 1998). It contains two domains, MW = 32 and 23 kDa, which are linked by disulfide bridges (Komath *et al.*, 1996). Chemical-modification studies indicated the presence of histidine residues in the sugar-binding site of SGSL (Komath *et al.*, 1998). The lectin has binding sites for porphyrins which appear to be different from those for carbohydrates (Komath *et al.*, 2000). As a first step in the detailed structure analysis of SGSL, we report here its crystallization and structure solution using molecular replacement and partial refinement of the structure.

2. Materials and methods

SGSL was isolated and purified as described previously (Komath *et al.*, 1996; Komath & Swamy, 1998). Crystals of two forms of SGSL, form I and form II, were grown in the presence of lactose and methyl-α-D-galactose, respec-

Table 1
Crystal data and data-collection statistics.

Values in parentheses refer to the last resolution shell.

	Form I	Form II	
		Low temp. (100 K)	Room temp. (298 K)
Space group	$P622$	$P6_122$	$P6_122$
Unit-cell parameters			
a (Å)	118.9	98.9	102.1
b (Å)	118.9	98.9	102.1
c (Å)	228.4	266.9	273.1
No. of subunits/asym. unit	1	1	1
Solvent content (%)	67	59	62
Data collection			
Resolution (Å)	6.0	3.0 (3.1–3.0)	3.1 (3.2–3.1)
No. of observations	—	65920	51684
No. of unique reflections	—	15685	14799
Multiplicity	—	4.2 (3.7)	3.5 (3.3)
Completeness (%)	—	98.0 (92.3)	92.7 (92.2)
$R_{\rm merge}$† (%)	—	12.1 (38.2)	16.0 (56.3)
Number of reflections with $I > 2\sigma(I)$ (%)	—	85.9 (65.1)	64.8 (39.1)

† $R_{\rm merge} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle$.

tively. Crystal form I was obtained by the hanging-drop method by equilibrating a 10 µl drop of 10 mg ml^{-1} protein in the presence of 4 m*M* lactose in 0.02 *M* phosphate buffer pH 7.0 containing 0.1 *M* NaCl and 0.025%(*w*/*v*) sodium azide against a reservoir solution of 80% saturated ammonium sulfate in the same buffer. The crystals grew to final dimensions of 0.4 × 0.2 × 0.08 mm in 60 d. Crystals of form II were also grown similarly (Fig. 1). The drop contained 10 µl of 40 mg ml^{-1} protein in the presence of 10 m*M* methyl-α-D-galactose, 5 m*M* β-mercaptoethanol and 1 µl of 30% PEG 400. The reservoir solution contained 1 ml of 80% saturated ammonium sulfate in the same buffer. These crystals grew to maximum dimensions of 1.5 × 1.0 × 0.3 mm in about 45 d. Crystal form I diffracted to about 6 Å, while form II diffracted to 3 Å resolution. Diffraction data from form II were collected at room temperature (298 K) and at low temperature (100 K) using flash-freezing with a MAR Research imaging plate mounted on a Rigaku RU-200 X-ray generator. The data were processed and scaled using the *MAR-XDS* (Kabsch, 1988) and *DENZO* (Otwinowski, 1993) programs. *AMoRe* (Navaza, 1994) was used for molecular-replacement calculations, while preliminary refinement and model building were carried out using the programs *CNS* (Brunger *et al.*, 1998) and *O* (Jones *et al.*, 1991), respectively.

Figure 1
Crystals of snake gourd seed lectin (form II).

3. Results and discussion

Crystal data and details of data collection are given in Table 1. The solvent contents of both forms, assuming one molecule in the asymmetric unit, are within the range normally observed for protein crystals (Matthews, 1968).

Unlike the case of type II RIPs, the amino-acid sequences of type I RIPs from seeds of *Momordica charantia* and *Trichosanthes kirilowii* belonging to the Cucurbitaceae family are available. They exhibit considerable sequence similarity to each other and with well studied type I RIPs from the Cucurbitaceae family such as bryodin and luffin and the *A* chain of type II RIPs such as abrin, ricin and mistletoe lectin. Therefore, it appeared reasonable to expect the *A* chain of SGSL also to be homologous to the type II RIPs mentioned above. The recent preliminary X-ray studies and molecular-replacement calculation of the lectin from *T. kirilowii* strengthened this expectation (Li *et al.*, 2000). In confirmation, a molecular-replacement calculation with abrin-a (Tahirov *et al.*, 1995; PDB code 1abr) as the search model using the low-temperature data from form II led to satisfactory structure solution. A long loop (41–50) in the *A* chain and ten residues at the N-terminus of the *B* chain were omitted from the search model. The best solution had a correlation coefficient (CC) of 0.345 and an *R* factor of 0.489 and led to satisfactory crystal packing. The corresponding next best solution had CC = 0.289 and *R* = 0.504. Similar results were obtained when the ricin molecule (PDB code 1aai) was used as the search model. Independent calculations using the *A* chain and *B* chain as search models also led to the same solution. A Fourier map calculated after rigid-body refinement using a polyalanine model showed contiguous density for most of the polypeptide chains. This was particularly so in the case of the *A* chain.

Further refinement of the structure was hampered by the absence of sequence information. However, it was found that the *A* chain of type II RIPs such as ricin and abrin and type I RIPs from the Cucurbitaceae family exhibited considerable sequence homology. Thus, it appeared that the sequence of the *A* chain is substantially conserved across families. Sequence identity among the Cucurbitaceae type I RIP sequences is as high as 52%; there is also a 38% sequence identity with ricin and abrin sequences. Careful examination of these sequences and successive electron-density maps led to the identification of 31% of the side chains in the *A* chain. This could not be performed with the *B* chain in the absence of the availability of homologous sequences from the same family. Refinement of the model using data in the resolution range 20–

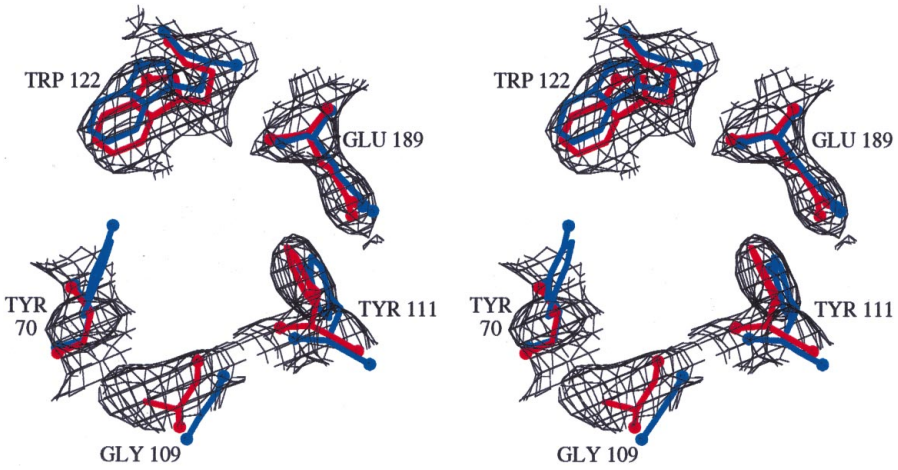

Figure 2
Electron density in the putative active site in the *A* domain of the lectin. The relevant residues in the lectin and in trichosanthin are shown in red and blue, respectively. The numbering corresponds to that in trichosanthin. No density corresponding to the side chain of Tyr70 in trichosanthin is seen in the map. Density for a side chain is present corresponding to Gly109.

3.0 Å employing the 'mlf' target was performed, resulting in an R factor of 0.34 and an $R_{\rm free}$ of 0.40 (with 8% omitted data). Further refinement of the structure is in progress. Sequence determination using chemical means is also under way. However, the preliminary crystallographic results and the available biochemical data indicate that SGSL is homologous to type II RIPs. Our studies also provide a rationale for the structure not possessing the same level of ribosome-inactivating activity as abrin-a and ricin. The crystal structure of trichosanthin (TCS) complexed with a substrate analogue revealed the role of conserved amino-acid residues in the active-site cleft in the N-glycosidase activity of RIPs (Xiong *et al.*, 1994). A comparison of our preliminary SGSL model with that of RIPs indicated that the fully conserved Tyr70 (TCS numbering) that is important for the biological activity is replaced by a non-aromatic residue in SGSL (Fig. 2). Also, a conserved Gly109 appears to be replaced by a larger residue. A similar result was also obtained in preliminary structural studies of *Trichosanthes kirilowii* lectin 1 (Li *et al.*, 2000). A detailed analysis of the structure of SGSL should thus reveal the structural basis for the difference in biological activity between SGSL and other RIPs.

The intensity data were collected at the X-ray Facility for Structural Biology at the Institute, supported by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT). Facilities at the Supercomputer Education and Research Centre and the Interactive Graphics based facility and Distributed Information Centre (both supported by DBT) were used in the work. The work at Bangalore and Hyderabad was supported by Research Grants from the DST to MV and the Council of Scientific and Industrial Research (CSIR) to MJS.

References

Bostwick, D. E., Skaggs, M. I. & Thompson, G. A. (1994). *Plant Mol. Biol.* **26**, 887–897.
Bouckaert, J., Hamelryck, T., Wyns, L. & Loris, R. (1999). *Curr. Opin. Struct. Biol.* **9**, 572–577.
Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Falasca, A. I., Abbondanza, A., Berbieri, A., Bolognesi, A., Rossi, C. A. & Stripe, A. (1989). *FEBS Lett.* **246**, 159–162.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
Komath, S. S., Kenoth, R., Giribabu, L., Maiya, B. G. & Swamy, M. J. (2000). *J. Photochem. Photobiol. B*, **55**, 49–55.
Komath, S. S., Nadimpalli, S. K. & Swamy, M. J. (1996). *Biochem. Mol. Biol. Int.* **39**, 243–252.
Komath, S. S., Nadimpalli, S. K. & Swamy, M. J. (1998). *Biochem. Mol. Biol. Int.* **44**, 107–116.
Komath, S. S. & Swamy, M. J. (1998). *Curr. Sci.* **75**, 608–611.
Krauspenhaar, R., Eschenburg, S., Perbandt, M., Kornilov, V., Konareva, N., Mikailova, I., Stoeva, S., Wacker, R., Maier, T., Singh, T., Mikhailov, A., Voelter, W. & Betzel, C. (1999). *Biochem. Biophys. Res. Commun.* **257**, 418–424.
Li, M., Wang, Y.-P., Chai, J.-J., Wang, K.-Y. & Bi, C. (2000). *Acta Cryst.* D**56**, 1073–1075.
Lis, H. & Sharon, N. (1998). *Chem. Rev.* **98**, 637–674.
Loris, R., Hamelryck, T., Bouckert, J. & Wyns, L. (1998). *Biochim. Biophys. Acta*, **1383**, 9–36.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. (1993). *DENZO. An Oscillation Data Processing Program for Macromolecular Crystallography*. Yale University, New Haven, CT, USA.
Padma, P., Komath, S. S., Nadimpalli, S. & Swamy, M. J. (1999). *Phytochemistry*, **50**, 363–371.
Ren, J., Wang, Y., Dong, Y. & Stuart, D. I. (1994). *Structure*, **2**, 7–16.
Rutenber, E., Katzin, B. J., Ernst, S., Collins, E. J., Misna, D., Ready, M. P. & Robertus, J. D. (1991). *Proteins Struct. Funct. Genet.* **10**, 240–250.
Sairam, M. R., Srinivasa, B. R. & Marcil, J. (1993). *Biochem. Mol. Biol. Int.* **31**, 571–581.
Singh, V. & Kar, S. K. (1992). *Ind. J. Biochem. Biophys.* **29**, 31–41.
Tahirov, T. H., Lu, T., Liaw, Y., Chen, Y. & Lin, J. (1995). *J. Mol. Biol.* **250**, 354–367.
Transue, T. R., Smith, A. K., Mo, H., Goldstein, I. J. & Saper, M. A. (1997). *Nature Struct. Biol.* **4**, 779–783.
Vijayan, M. & Chandra, N. (1999). *Curr. Opin. Struct. Biol.* **9**, 707–714.
Wang, H. & Ng, T. B. (1998). *Biochem. Biophys. Res. Commun.* **253**, 143–146.
Xiong, J. P., Xia, Z. X. & Wang, Y. (1994). *Nature Struct. Biol.* **1**, 695–700.